SHIFTED BARRIER METHODS FOR LINEAR PROGRAMMING

Philip E. GILL
Department of Mathematics
University of California, San Diego
La Jolla, California 92093

Walter MURRAY, Michael A. SAUNDERS
Systems Optimization Laboratory, Department of Operations Research
Stanford University
Stanford, California 94305

Margaret H. WRIGHT
AT&T Bell Laboratories
Murray Hill, New Jersey 07974

Technical Report SOL 87-9*
Reprinted 1991

Presented at the Second Conference on
Progress in Mathematical Programming,
Asilomar, California, March 1–4, 1987

Abstract

Almost all interior-point methods for linear programming published to date can be viewed as application of Newton's method to minimize (approximately) a sequence of logarithmic barrier functions. These methods have been observed to suffer from certain undesirable features—most notably, badly behaved subproblems arising from large derivatives of the barrier function near the constraint boundary. We define a shifted barrier method intended to avoid these difficulties. Convergence proofs are given for both exact and approximate solution of the shifted barrier subproblems corresponding to the primal and dual problems. We also show that the usual Newton iterates eventually satisfy the requirements for sufficiently accurate approximate solutions. Several practical strategies are suggested for initializing and updating the weights and shifts.

1. Introduction

Consider the inequality-constrained optimization problem

$$\underset{x\in\mathbb{R}^n}{\text{minimize}}\ \ f(x) \quad \text{subject to} \quad c(x) \geq 0, \tag{NIP}$$

*The material contained in this report is based upon research supported by the Air Force Office of Scientific Research Grant 87-01962; the U.S. Department of Energy Grant DE-FG03-87ER25030; National Science Foundation Grant CCR-8413211; and the Office of Naval Research Contract N00014-87-K-0142.

where $c(x)$ is an l-vector of constraint functions.

A "classical" approach (dating from the 1950's) for solving (NIP) is to create a *barrier function*, which is a weighted combination of $\varphi(x)$ and a barrier term for each constraint. (A barrier term is a differentiable function with a positive singularity at the constraint boundary.) We treat only the logarithmic barrier function, first proposed by Frisch [Fri55], in which the barrier term corresponding to the constraint $c_i(x) \geq 0$ is $-\ln c_i(x)$, so that the barrier function associated with (NIP) is

$$B(w, x) \equiv \varphi(x) - \sum_{i=1}^{l} w_i \ln c_i(x), \tag{1.1}$$

where $w_i > 0$ for all i. When all $\{w_i\}$ have a common value (say, μ), μ is usually called the *barrier parameter*.

Let $x^*(w)$ denote an unconstrained minimizer of $B(w, x)$. Under mild conditions on φ and $\{c_i\}$, it can be shown that

$$\lim_{\|w\| \to 0} x^*(w) = x^*,$$

where x^* is a solution of the original problem (NIP). In a *practical* barrier-function method, the requirement of exact minimization is relaxed, and approximate minimizers of $B(w^k, x)$ are computed for a sequence of weight vectors $\{w^k\}$ converging to zero. Since the barrier term is undefined outside the feasible region, a strictly feasible starting point (satisfying $c(x) > 0$) is required for each minimization, and only strictly feasible iterates are generated.

A complementary and similarly longstanding approach to solving problem (NIP) involves definition of a *penalty function*, a weighted combination of φ and a penalty term (a continuous function that measures constraint violations). The most popular penalty term has traditionally been the squared two-norm, which gives the *quadratic* penalty function. Under mild conditions, the unconstrained minimizers of a sequence of quadratic penalty functions converge to x^* if the weight corresponding to each violated constraint becomes infinite.

For a complete discussion of barrier methods, see Fiacco [Fia79]. Classical barrier and penalty methods are described in Fiacco and McCormick [FM68]. Fletcher [Fle81] and Gill, Murray and Wright [GMW81] give overviews of barrier and penalty methods.

Powell [Pow69] derived the well known class of augmented Lagrangian methods from the idea of adaptively *shifting* the constraint boundaries in a quadratic penalty method, i.e., replacing the constraint $c_i(x) \geq 0$ by $c_i(x) + s_i \geq 0$, where the nonnegative shift s_i may change as the algorithm proceeds. With proper choice of shifts and weights, the penalty function is equivalent to an augmented Lagrangian function, and the sequence of penalty-function minimizers converges to x^* for a *finite* weight vector. An obvious extension to barrier methods is to shift the location of the singularity. (A brief description of such an algorithm in the context of *active-set* methods for nonlinear programming is given by Osborne [Osb72].) However, there is no clear relationship between a shifted barrier function and the Lagrangian function, and such methods have not been widely used.

It is now generally recognized that essentially all interior-point methods for linear programming inspired by Karmarkar's [Kar84] projective method are closely related to application of Newton's method to a sequence of barrier functions. Newton's method is based on minimizing a local quadratic model of the barrier function derived from first and second derivative information at the current iterate. Unfortunately, several difficulties can arise because of the nature of barrier functions. The extreme nonlinearity of the barrier term near the boundary means that a quadratic model may be accurate only in a very small neighborhood of the current point. For a degenerate linear program, the Hessian of the barrier function becomes increasingly ill-conditioned at $x^*(w)$ when $\|w\|$ is very small (see Section 4). Finally, a strictly interior starting point may be inconvenient or impossible to obtain.

The purpose of this paper is to derive shifted barrier methods for linear programming specifically designed to avoid these difficulties. In Section 2, we describe a shifted barrier method for a linear program in standard form, and give a relationship between the weights and shifts that ensures convergence. A shifted barrier method for the dual formulation of a linear program is given in Section 3. In Section 4, we describe the behavior of shifted barrier subproblems in the neighborhood of the solution. The properties of Newton's method applied to both the primal and dual subproblems are discussed in Section 5. Finally, methods for initializing and updating the shifts are proposed in Section 6, which concludes with the definition of a *Lagrangian* shifted barrier method for both primal and dual problems.

1.1. Notation and background

The *primal linear program* is taken to be of the form

$$\text{LP} \qquad \begin{array}{ll} \underset{x}{\text{minimize}} & c^T x \\ \text{subject to} & Ax = b, \qquad x \ge 0, \end{array} \tag{1.2}$$

where A is an $m \times n$ matrix of full rank with $m \le n$, and both b and c are nonzero. We assume throughout that a solution exists.

The *dual linear program* associated with (1.2) is given by

$$\text{DLP} \qquad \begin{array}{ll} \underset{\pi\in\mathbb{R}^m}{\text{minimize}} & -b^T \pi \\ \text{subject to} & -A^T\pi \ge -c, \end{array} \tag{1.3}$$

where for consistency we have posed the problem in terms of minimization and lower-bound constraints.

The feasible point x^* solves the primal problem (1.2) if an m-vector π^* exists such that the n-vector z^* defined by

$$z^* \equiv c - A^T\pi^* \tag{1.4}$$

satisfies

$$z^* \ge 0 \ \text{ and} \tag{1.5a}$$

$$z_j^* x_j^* = 0, \quad j = 1, \ldots, n. \tag{1.5b}$$

The vectors π^* and z^* are the Lagrange multipliers associated with the constraints $Ax = b$ and $x \geq 0$ respectively, and π^* is a solution of the dual problem. Let $\mathcal{X}^*$ denote the (nonempty) set of all solutions of LP, and Π^* denote the set of all solutions of DLP. Associated with each $\pi^* \in \Pi^*$ is z^*. Let Z^* denote the set of all z^*. It will be convenient to introduce the notation x^{*k}, π^{*k} and z^{*k} to denote the nearest element in $\mathcal{X}^*$, Π^* and z^* to x_k, π^k and z^k respectively. We shall sometimes need to refer to nondegenerate points. A point is said to be nondegenerate if the number of constraints that are exactly satisfied at the point is not greater than the number of variables. In the primal space this implies a point x is nondegenerate if no more than $n - m$ elements of x are zero. In the dual space it implies no more than m elements of z are zero.

2. Shifted Barrier Methods for the Primal Problem

Throughout this paper, s and w will denote vectors called *shifts* and *weights*. The *shifted barrier function* corresponding to problem (NIP) is

$$B(w, s, x) = \varphi(x) - \sum_{i=1}^{l} w_i \ln(c_i(x) + s_i), \tag{2.1}$$

where $w_i > 0$ and $s_i \geq 0$. Any unconstrained minimizer of (2.1) lies strictly interior with respect to the shifted constraints. Choosing $s_i = 0$ for all i gives an "unshifted" barrier function.

Specializing this definition to linear programming, the *primal shifted barrier subproblem* associated with problem (1.2) is

$$\text{SBP}(w, s) \qquad \begin{array}{ll} \underset{x}{\text{minimize}} & F_P(x) \equiv c^T x - \displaystyle\sum_{j=1}^{n} w_j \ln(x_j + s_j) \\ \text{subject to} & Ax = b. \end{array} \tag{2.2}$$

We shall show the explicit dependence of SBP on w and s only when necessary.

Throughout this section, we assume that the constraints of problem LP define a bounded feasible region. This assumption guarantees that the feasible region for the dual has a nontrivial interior. Any solution of SBP must satisfy $Ax = b$. Since the feasible region of LP is bounded and the objective function of SBP is strictly convex, its solution, denoted by $x^*(w, s)$, is unique. Since A has full rank, the associated Lagrange multiplier vector $\pi^*(w, s)$ associated with the equality constraints of SBP is also unique, and $\|\pi^*(w, s)\|$ is bounded (see Stewart [Ste88]).

2.1. The role of the Lagrangian function

The *Lagrangian function* associated with SBP is

$$L(x, \pi, w, s) = c^T x - \sum_{j=1}^{n} w_j \ln(x_j + s_j) - \pi^T (Ax - b), \tag{2.3}$$

where π is usually interpreted as a vector of Lagrange multipliers associated with the constraints $Ax = b$. For any m-vector π, it is convenient to associate an n-vector z defined by

$$z \equiv c - A^T \pi. \tag{2.4}$$

The gradient of the Lagrangian (2.3) with respect to x is denoted by g_L, and is given by

$$g_L(x, \pi) \equiv \nabla_x L(x, \pi) = c - \sum_{j=1}^{n} \frac{w_j}{x_j + s_j} e_j - A^T \pi \tag{2.5a}$$

$$= z - \sum_{j=1}^{n} \frac{w_j}{x_j + s_j} e_j, \tag{2.5b}$$

where e_j denotes the j-th coordinate vector. (As with SBP, the dependence of g_L on w and s is suppressed if the meaning is clear.) Because of the properties of SBP, the Lagrangian (2.3) has a unique stationary point at $(x^*(w, s),\ \pi^*(w, s))$.

In an unshifted barrier method, $\|w\|$ must converge to zero in order for minimizers of the barrier function (1.1) to converge to the solution of the original problem. We now give a more relaxed condition that ensures convergence of the sequence of solutions of the shifted barrier subproblem when there exists a dual solution that is a nondegenerate point.

Consider a sequence of shifted barrier subproblems $\{\text{SBP}(w^k, s^k)\}$, where w^k and s^k denote the vectors of weights and shifts defining the k-th subproblem. Let the solution of the k-th subproblem be x^k, with π^k the corresponding Lagrange multiplier.

Lemma 2.1. *Let $\{w^k\}$ and $\{s^k\}$ be sequences of positive weights and shifts such that*

$$\lim_{k\to\infty} \frac{w_j^k}{s_j^k} = z_j^*, \tag{2.6}$$

where z^* is defined by (1.4) for any $\pi^* \in \Pi^*$. If there exists a dual solution that is a nondegenerate point of the dual of LP, then $\lim_{k\to\infty} x^k = x^*$.

Proof. The existence of a nondegenerate dual solution implies x^* is unique. Convergence of the ratio $\{w_j^k / s_j^k\}$ to z_j^* implies that $w_j^k / s_j^k = z_j^* + \epsilon_j^k$, where $\epsilon_j^k \to 0$. (We sometimes use the notation $\tau^k \to \gamma$ to mean $\lim_{k\to\infty} \tau^k = \gamma$.) Because of the optimality conditions for LP (cf.. (1.5b)), we know that $z_j^* x_j^* = 0$. Accordingly, w_j^k may be written as

$$w_j^k = s_j^k z_j^* + s_j^k \epsilon_j^k = (x_j^* + s_j^k) z_j^* + s_j^k \epsilon_j^k,$$

which gives

$$\frac{w_j^k}{x_j^* + s_j^k} = z_j^* + \frac{s_j^k \epsilon_j^k}{x_j^* + s_j^k}. \tag{2.7}$$

Let L^k denote the Lagrangian function (2.3) defined with weights w^k and shifts s^k. From (2.5), the gradient of L^k with respect to x evaluated at (x^*, π^*) is given by

$$g_L^k(x^*, \pi^*) = c - \sum_{j=1}^{n} \frac{w_j^k}{x_j^* + s_j^k} e_j - A^T \pi^*.$$

Substituting from (2.7) gives

$$g_L^k(x^*, \pi^*) = c - z^* - \sum_{j=1}^{n} \frac{s_j^k \epsilon_j^k}{x_j^* + s_j^k} e_j - A^T \pi^* = -\sum_{j=1}^{n} \frac{s_j^k \epsilon_j^k}{x_j^* + s_j^k} e_j.$$

Because $x^* \geq 0$ and $s^k > 0$, the quotient $s_j^k/(x_j^* + s_j^k)$ is uniformly bounded, and consequently $g_L^k(x^*, \pi^*) = O(\|\epsilon^k\|)$. Since the solution of SBP is unique, convergence of ϵ^k to 0 and boundedness of the feasible region of (1.2) imply that $\lim_{k\to\infty} x^k = x^*$, as required. ∎

2.2. Convergence of the primal shifted barrier method

There are many ways in which the shifts may be defined. However, some restrictions are necessary to ensure that the sequence $\{x_j^k\}$ is feasible in the limit. We shall define one particular algorithm for s_j^k.

Those chosen were done with the view to the knowledge that in solving the $(k+1)$-th subproblem we may wish to use as an initial point the solution of the k-th subproblem. This being the case we require $x^k + s^{k+1} > 0$. It will be always possible to satisfy this condition if $s^{k+1} < s^k$ only if $x^k > 0$.

The algorithm requires three preassigned scalars: ν, ρ and θ. The value of ν $(0 < \nu < 1)$ determines the maximum potential reduction in any shift at each iteration. The values of θ $(0 < \theta < \infty)$ and ρ $(0 < \rho < \infty)$ are used to limit the oscillations in x_j for components whose reduced costs are going to zero.

Given an initial estimates $\{x_j^0\}$ and $\{\pi_i^0\}$ of the primal and dual variables, the initial shifts $\{s_j^1\}$ are any set of bounded positive numbers and the initial weights are $w_j^1 = z_j^0 s_j^1$, where $z^0 \equiv c - A^T\pi^0$. Thereafter, x^k denotes the (unique) minimizer of SBP(w^k, s^k) subject to $x + s^k > 0$, π^k denotes the associated multiplier vector and $z^k \equiv c - A^T\pi^k$.

Associated with each variable are two numbers μ_j and a_j. The scalar μ_j holds the smallest reduced cost computed so far. The scalar a_j defines a threshold value that is used to test if the z_j-s appear to be converging to zero. A decreasing reduced cost that is less than a_j triggers a reduction in the shift. Initially, we set $\mu_j = z_j^0$ and $a_j = \nu z_j^0$.

Algorithm WS (Update for the weights and shifts)

$\sigma_1 \leftarrow s_j^k; \quad \sigma_2 \leftarrow s_j^k;$
$\mu_j \leftarrow \min\{\mu_j, z_j^k\};$
if $x_j^k < -\theta/k$ **then**
  $n_j \leftarrow n_j + 1;$
else if $x_j^k > \rho\mu_j$ **and** $n_j > 0$ **then**
  $\sigma_1 \leftarrow \nu\rho\mu_j; \quad n_j \leftarrow 0;$
end if
if $z_j^k < z_j^{k-1}$ **and** $z_j^k < a_j$ **then**
  $\sigma_2 \leftarrow \nu s_j^k; \quad a_j \leftarrow \nu a_j;$
end if
$s_j^{k+1} \leftarrow \min\{\sigma_1, \sigma_2\}; \quad w_j^{k+1} \leftarrow z_j^k s_j^{k+1};$

The following lemma summarizes the properties that are needed to prove the main convergence theorem. Any alternative definition of the shifts must give a sequence $\{s_j^k\}$ that satisfies the same properties.

Lemma 2.2. *If the shifts are defined by Algorithm WS, then the sequence* $\{s_j^k\}$ *has the following properties.*

1. *If* $\{k_i\}$ *is a subsequence such that* $\lim_{i\to\infty} z_j^{k_i} = 0$*, then* $\lim_{k\to\infty} s_j^k = 0$.
2. *There exists a bounded positive constant* M *such that* $0 \le s_j^k \le M$ *for all* $k > 0$ *and* $M s_j^k > s_j^i$ *for all* $i > k$.

In the main theorem, we establish convergence without any assumptions regarding the degeneracy of either LP or its dual.

Theorem 2.1. *(Convergence with exact solutions of SBP.) Let* $\{w^k\}$ *and* $\{s^k\}$ *denote the weights and shifts generated by Algorithm WS. Let* x^k *denote the (unique) minimizer of SBP*(w^k, s^k) *subject to* $x + s^k > 0$, π^k *the associated multiplier vector and* $z^k \equiv c - A^T\pi^k$*. Then*

$$\lim_{k\to\infty} \|x^{*k} - x^k\| = 0 \;\; \textit{and} \;\; \lim_{k\to\infty} \|\pi^{*k} - \pi^k\| = 0.$$

Proof. First, we use an inductive argument to show that the elements of w^k and z^k are positive for all k. Assume that $w^l > 0$ for $l \ge 0$. By definition, g_L (2.5) vanishes at (x^l, π^l), which implies that

$$z_j^l = c_j - a_j^T\pi^l = \frac{w_j^l}{x_j^l + s_j^l}. \tag{2.8}$$

Since x^l is interior with respect to the shifted constraints, we have $x_j^l + s_j^l > 0$, and (2.8) then implies that $z_j^l > 0$. The positivity of w^{l+1} follows directly from the definition $w_j^{l+1} = z_j^l s_j^{l+1}$ (see Algorithm WS). Since w^0 is positive by construction,

$w^l > 0$ for all l. Furthermore, the positivity of z^l implies that π^l is feasible for the dual problem (1.3).

Premultiplying the relation $z^k = c - A^T\pi^k$ by $(x^k)^T$ gives

$$(x^k)^T z^k = (x^k)^T(c - A^T\pi^k).$$

Since x^k satisfies $Ax^k = b$, we have

$$(x^k)^T z^k = c^T x^k - b^T\pi^k. \tag{2.9}$$

A similar argument with z^{k-1} gives

$$(x^k)^T z^{k-1} = c^T x^k - b^T\pi^{k-1}. \tag{2.10}$$

Combining (2.9) and (2.10), we obtain

$$b^T\pi^k - b^T\pi^{k-1} = (x^k)^T(z^{k-1} - z^k). \tag{2.11}$$

Next we substitute $w_j^k = z_j^{k-1} s_j^k$ in (2.8) to obtain

$$z_j^k = \frac{w_j^k}{x_j^k + s_j^k} = \frac{z_j^{k-1} s_j^k}{x_j^k + s_j^k}, \quad \text{or} \quad z_j^{k-1} = \frac{z_j^k(x_j^k + s_j^k)}{s_j^k}. \tag{2.12}$$

Using the latter expression to substitute for z_j^{k-1} in (2.11) gives

$$b^T\pi^k - b^T\pi^{k-1} = \sum_{j=1}^{n} \frac{z_j^k (x_j^k)^2}{s_j^k} > 0. \tag{2.13}$$

Since π^k is feasible for the dual problem (1.3), it follows that $b^T\pi^k \le b^T\pi^* < \infty$. The sequence $\{b^T\pi^k\}$ is monotonically increasing and bounded above, and therefore converges, i.e., $\lim_{k\to\infty}(b^T\pi^k - b^T\pi^{k-1}) = 0$. It follows from (2.13) that

$$\lim_{k\to\infty} \sum_{j=1}^{n} \frac{z_j^k (x_j^k)^2}{s_j^k} = 0.$$

Since $\|s^k\|$ is bounded, we have

$$z_j^k x_j^k \to 0 \;\text{ and }\; (z^k)^T x^k \to 0. \tag{2.14}$$

Relations (2.9) and (2.14) then give

$$\lim_{k\to\infty}(c^T x^k - b^T\pi^k) = 0, \;\text{ or equivalently, }\; \lim_{k\to\infty} c^T x^k = \lim_{k\to\infty} b^T\pi^k. \tag{2.15}$$

The standard duality connection between LP and DLP implies that

$$c^T x^* = b^T\pi^*, \tag{2.16}$$

so that (2.15) gives

$$\lim_{k\to\infty} c^T x^k = \lim_{k\to\infty} b^T \pi^k \le b^T \pi^* = c^T x^*, \tag{2.17}$$

which leads to

$$\lim_{k\to\infty} c^T x^k \le c^T x^*. \tag{2.18}$$

We shall now show a subsequence $\{x^{k_j}\}$ exists such that

$$\lim_{j\to\infty} \|x^{k_j} - \bar{x}^j\| = 0,$$

where $\bar{x}^j$ is the closest feasible point to x^{k_j}. The sequence $\{x_j^k\}$ must behave as one of the following cases:

(1) $\liminf_{k\to\infty} x_j^k \ge 0$;

(2) $\liminf_{k\to\infty} x_j^k < 0$ and $\limsup_{k\to\infty} x_j^k < 0$;

(3) $\liminf_{k\to\infty} x_j^k < 0$ and $\limsup_{k\to\infty} x_j^k > 0$; or

(4) $\liminf_{k\to\infty} x_j^k < 0$ and $\limsup_{k\to\infty} x_j^k = 0$.

If case (1) occurs, the component of x^k is feasible. If case (2) occurs, then for some $k > K$ we have $x_i^k < -\theta < 0$. Given the identity

$$z_j^k = \frac{s_j^k}{x_j^k + s_j^k} z_j^{k-1},$$

it follows that z_j^k is monotonically increasing for $k > K$, which contradicts the result that $\lim_{k\to\infty} z_j^k = 0$ implied by (2.14).

In case (3), there must exist a subsequence, say $\{x_j^{k_i}\}$ such that $x_j^{k_i} < -\theta < 0$. It follows from (2.14) that $\lim_{i\to\infty} z_j^{k_i} = 0$. Hence from condition (c) on the shifts we have $\lim_{i\to\infty} s_j^{k_i} = 0$. This contradicts the condition that $x_j^{k_i} + s_j^{k_i} > 0$.

It follows from lemma (6.3) that if even if case (4) occurs for all elements of x^k, there exists a subsequence such that

$$\lim_{i\to\infty} \|\widehat{x}^{l_i} - \widehat{x}^i\| = 0,$$

where $\widehat{x}^k$ denotes those elements of x^k for which case (4) occurs. We have shown that the convergence behavior is described by case (1), so that

$$\lim_{i\to\infty} \|x^{k_i} - \bar{x}^i\| = 0.$$

It follows immediately from this result and (2.17) that

$$\lim_{k\to\infty} c^T x^k = \lim_{k\to\infty} b^T \pi^k = b^T \pi^* = c^T x^*. \tag{2.19}$$

Finally, since π^k is dual feasible we must have

$$\lim_{k\to\infty} \|\pi^{*k} - \pi^k\| = 0$$

and

$$\lim_{k\to\infty} \|z^{*k} - z^k\| = 0. \tag{2.20}$$

Note we have yet to use condition (b). It follows from (2.19) that in order to prove the final result, it is necessary only to show that $\lim_{k\to\infty} d_k = 0$, where d_k is the shortest distance from x^k to the feasible region. Suppose there exists a subsequence such that $x_j^{k_i} \le -\epsilon < 0$. It follows that $\lim_{i\to\infty} z_j^{k_i} = 0$. From condition (b) we obtain $\lim_{i\to\infty} s_j^{k_i} = 0$, which implies that for i sufficiently large, $x_j^{k_i} + s_j^{k_i} < 0$. Since this contradicts our assumptions regarding the definition of the shifts, it follows no such subsequence exists and

$$\lim_{k\to\infty} \|x^{*k} - x^k\| = 0.$$

∎

It should be noted that there is no difficulty in constructing a sequence such that $\lim_{k\to\infty} s_j^k = 0$ if $\lim_{i\to\infty} z_j^{k_i} = 0$ since this latter limit implies the existence of a subsequence $\{l_i\}$ such that $x_j^{l_i} > 0$.

Corollary 2.1. *If condition (b) in Theorem 2.1 is replaced by the assumption that Z^* is a set of nondegenerate points then the results of the theorem still hold.*

Proof. Since condition (b) was not used to show convergence of $\{\pi^k\}$ we need only be concerned with showing $\{x^k\}$ converges to a feasible point. If z^* is a nondegenerate point then x^* is unique. Consider the following linear program:

$$\begin{array}{lll} \mathrm{LP}^k & \underset{x}{\text{minimize}} & c^T x \\ & \text{subject to} & Ax = b, \\ & & x_j \ge 0,\ j \notin \mathcal{J}^k, \qquad x_j \ge x_j^k,\ j \in \mathcal{J}^k, \end{array} \tag{2.21}$$

where $\mathcal{J}^k$ is the set of indices such that $x_j^k < 0$. Let $\widehat{X}^k$ denote the set of solutions of LP^k and $(\widehat{Z}^k, \widehat{\Pi}^k)$ the corresponding dual solutions. Clearly x^k is feasible for LP^k and (z^k, π^k) is dual feasible. It follows from (2.14) that

$$\lim_{k\to\infty} \sum_{i\notin\mathcal{J}^k} x_i^k z_i^k = 0.$$

Consequently, $\lim_{k\to\infty} \|z^k - \hat{z}^k\| = 0$ and $\lim_{k\to\infty} \|\pi^k - \hat{\pi}^k\| = 0$, where $\hat{z}^k$ and $\hat{\pi}^k$ are the nearest elements in $\widehat{Z}^k$ and $\widehat{\Pi}^k$ to z^k and π^k respectively. This implies that in the limit there exists a dual solution of LP^k that is a nondegenerate point. Consequently, in the limit the primal solution of LP^k is unique. From (2.19) we have

$\lim_{k\to\infty} c^T\bar{x}^k = c^T x^*$. We can, therefore, infer that the solution of LP^k is unique only if $\lim_{k\to\infty} x^k = x^*$. ∎

It follows from

$$\lim_{k\to\infty} \sum_{i\in\mathcal{J}^k} x_i^k z_i^k + b^T\pi^k = b^T\pi^*$$

that in the limit all dual solutions of LP^k are dual solutions of LP. Consequently all dual solutions of LP^k are nondegenerate points.

Corollary 2.2. *Under the same assumptions as in theorem 2.1 except condition (b) is replaced by the assumption that x^* and z^* are unique (this would be true if x^* and z^* are nondegenerate points of LP and the dual of LP respectively), then $\lim_{k\to\infty} x^k = x^*$ and $\lim_{k\to\infty} \pi^k = \pi^*$. Moreover, condition (c) does not restrict the choice of s^k.*

Proof. We need only show the choice of s^k is not restricted by condition (c) since $\lim_{k\to\infty} x^k = x^*$ and $\lim_{k\to\infty} z^k = z^*$ follows immediately from the corollary 2.1.

The shifts are only restricted if there exists a subsequence where s^k is restricted. Suppose such a subsequence exists for the i-th shift. Since the previous corollary establishes the convergence of x^k and z^k we must either have $\lim_{k\to\infty} x_i^k = 0$ or $\lim_{k\to\infty} x_i^k > 0$. If it is the first case then $\lim_{k\to\infty} z_i^k > 0$, which implies there exists θ such that $z_i^k \geq \theta$ and s^k is not restricted by the rule. If $\lim_{k\to\infty} x_i^k > 0$ there exists K such that for all $k > K$, $x_i^k > \theta > 0$. This contradicts the existence of the subsequence, which requires there exists a subsequence $\{k_j\}$ such that $x_i^{k_j} < 0$. ∎

Although these corollarys show it is not necessary to have some elements of s^k tend to zero it will be seen later that this may be desirable.

In practice, it is likely to be inefficient to solve each shifted barrier subproblem exactly. The next theorem gives conditions under which a sequence of approximate solutions of SBP converges to the desired solution of LP. In Section 5 we show that these conditions are eventually satisfied when Newton's method is applied to solve SBP.

Theorem 2.2. *(Convergence with approximate solutions of SBP.) Let, x^k be an approximate solution of SBP(w^k, s^k). Let the sequence $\{\pi^k\}$ represent corresponding approximate Lagrange multiplier vectors, with $z^k = c - A^T\pi^k$. Assume the weights and shifts satisfy the conditions given in Theorem 2.1. Let β be a scalar such that $0 < \beta < 1$. If x^k and π^k satisfy*

(a) $Ax^k = b$, $x^k + s^k > 0$;

(b) $b^T\pi^k - b^T\pi^{k-1} \geq \beta \sum_{j=1}^{n} \dfrac{z_j^{k-1}(x_j^k)^2}{x_j^k + s_j^k}$; *and*

(c) $|(g_L^k)_j| < \frac{1}{2} \dfrac{s_j^k z_j^{k-1}}{k!(x_j^k + s_j^k)}$;

then

$$\lim_{k\to\infty} \|x^{*k} - x^k\| = 0 \ \ \textit{and} \ \ \lim_{k\to\infty} \|\pi^{*k} - \pi^k\| = 0.$$

Proof. It should be clear from Theorem 2.1 that the conditions given above are satisfied by the exact solution of SBP(w^k, s^k). The proof is similar in structure to that of Theorem 2.1.

For any x and π, rearranging expression (2.5) gives

$$z_j = (g_L^k)_j + \frac{w_j^k}{x_j + s_j^k}.$$

Hence, for $k \geq 0$ (by convention we define $z_j^{-1} = w_j^0/s_j^0$)

$$z_j^k = (g_L^k)_j + \frac{s_j^k z_j^{k-1}}{x_j + s_j^k}.$$

It follows when condition (c) is satisfied that

$$z_j^k = \psi_k \frac{s_j^k z_j^{k-1}}{x_j + s_j^k},$$

where ψ_k is either $1 + 0.5/k!$ or $1 - 0.5/k!$ depending on the sign of $(g_L^k)_j$. In either case it follows from induction that $z_j^k > 0$ and $w_j^k > 0$. Hence, π^k is dual feasible and $b^T\pi^k < b^T\pi^* < \infty$.

The sequence $\{b^T\pi^k\}$ is monotonically increasing (from assumption (b)) and bounded above, and therefore converges, i.e., $\lim_{k\to\infty}(b^T\pi^k - b^T\pi^{k-1}) = 0$. Furthermore,

$$\lim_{k\to\infty} z_j^k x_j^k = 0, \ \text{ so that } \ \lim_{k\to\infty} (z^k)^T x^k = 0. \tag{2.22}$$

Assumption (a) implies that $(x^k)^T z^k = c^T x^k - b^T\pi^k$, which, combined with (2.22), gives $\lim_{k\to\infty} c^T x^k = \lim_{k\to\infty} b^T\pi^k$. The remainder of the proof is now identical to Theorem 2.1. The condition (c) also ensures that Lemma 6.3 still applies. ∎

Corollary 2.3. *If condition (b) in Theorem 2.2 (given in theorem 2.1) is replaced by the assumption that Z^* is a set of nondegenerate points then the results of the theorem still hold.*

Corollary 2.4. *Under the same assumptions as in theorem 2.2 except condition (b) on the shifts is replaced by the assumption that x^* and z^* are unique (this would be true if x^* and z^* are nondegenerate points of LP and its dual respectively), then $\lim_{k\to\infty} x^k = x^*$ and $\lim_{k\to\infty} z^k = z^*$. Moreover, condition (c) does not restrict the choice of s^k.*

3. Shifted Barrier Methods for the Dual Linear Program

Since all constraints of the dual problem DLP (1.3) are inequalities, applying a shifted barrier transformation to this problem leads to a purely unconstrained subproblem:

$$\text{DSBP}(w,s) \qquad \underset{\pi\in\mathbb{R}^m}{\text{minimize}} \quad F_D(\pi) \equiv -b^T\pi - \sum_{j=1}^{n} w_j \ln(c_j - a_j^T\pi + s_j). \tag{3.1}$$

We assume in this section that the feasible region for DLP is bounded. Since DBSP is then a strictly convex function defined on a bounded domain, its unconstrained minimizer $\pi^*(w,s)$ is unique.

Given π, z is defined as usual as $z = c - A^T\pi$. For particular vectors w, s and z, we define x from

$$x_j = \frac{w_j}{z_j + s_j}. \tag{3.2}$$

The choice of the notation x is deliberate, since the gradient of F_D is $Ax - b$, and x approximates the solution of the primal problem LP when π approximates the solution of the dual. When the gradient of F_D vanishes, x satisfies $x > 0$ and $Ax = b$.

Let $\{w^k\}$ and $\{s_k\}$ denote sequences of weights and shifts. The following lemma, stated without proof, is analogous to Lemma 2.1.

Lemma 3.1. *Let $\{w^k\}$ and $\{s^k\}$ be infinite sequences of positive weights and shifts such that $w_j^k/s_j^k \to x_j^*$. Assume that LP has at least one primal solution that is a nondegenerate point, and let π^k denote the unconstrained minimizer of DSBP(w^k, s^k). Then $\lim_{k\to\infty} \pi^k = \pi^*$.* ∎

Theorem 3.1. *(Convergence with exact solutions of DSBP.) Let $\{w^k\}$ and $\{s^k\}$ be infinite sequences of weights and shifts, where $w^0 > 0$, $\infty > M > s^k > 0$ and $Ms^k > s^j$ for $j > k$. Let π^k denote the unique minimizer of DSBP with weights w^k and shifts s^k, and let x^k be given by (3.2). If*

(i)
$$w_j^k = x_j^{k-1} s_j^k \ \ \textit{for } k \ge 1, \tag{3.3}$$

(ii) *If $\{k_i\}$ is a subsequence such that $\lim_{i\to\infty} x_j^{k_i} = 0$ then $\lim_{k\to\infty} s_j^k = 0$.*

(iii) *If a pair of indices exists, say r and t, with $t > r$ such that*

$$z_i^r < -\theta/r$$

and

$$z_i^t > \bar{M}\phi_i^t,$$

where $\infty > \bar{M} > 0$ and $\phi_i^t = \min\{x_i^k \mid \ k = 1, \dots, t\}$ then $s_i^j < \bar{M}\phi_i^t$ for $j \ge t$.

It follows if the above conditions are satisfied that

$$\lim_{k\to\infty} \|x^{*k} - x^k\| = 0$$

and

$$\lim_{k\to\infty} \|\pi^{*k} - \pi^k\| = 0.$$

Proof. The proof is almost identical to that of Theorem 2.1. Recall that x^{k-1} and x^k are feasible for LP. Multiplying z^k by $(x^{k-1})^T$ and $(x^k)^T$ gives

$$(x^{k-1})^T z^k = (x^{k-1})^T c - (x^{k-1})^T A^T \pi^k = (x^{k-1})^T c - b^T \pi^k \tag{3.4}$$

and

$$(x^k)^T z^k = (x^k)^T c - b^T \pi^k. \tag{3.5}$$

Eliminating $b^T\pi^k$ from (3.4) and (3.5) gives

$$c^T x^{k-1} - c^T x^k = (z^k)^T (x^{k-1} - x^k) = \sum_{j=1}^{n} z_j^k (x_j^{k-1} - x_j^k). \tag{3.6}$$

Substituting w_j^k (3.3) in the definition of x_j^k gives

$$x_j^k = \frac{w_j^k}{z_j^k + s_j^k} = \frac{x_j^{k-1} s_j^k}{z_j^k + s_j^k}.$$

Using this expression to substitute for x_j^{k-1} in (3.6) we obtain

$$c^T x^{k-1} - c^T x^k = \sum_{j=1}^{n} \frac{(z_j^k)^2 x_j^k}{s_j^k}.$$

It follows that $\lim_{k\to\infty} c^T(x^{k-1} - x^k) = 0$, and the remainder of the proof follows that of Theorem 2.1 with the roles of z and x interchanged. ∎

The following two results analogous to Corollaries 2.1 and 2.2 may also be proved in a straightforward fashion.

Corollary 3.1. *Under the same assumptions as in Theorem 3.1 except condition (ii), but with the assumption that X^* is a set of nondegenerate points then* $\lim_{k\to\infty} \pi^k = \pi^*$.

Corollary 3.2. *Under the same assumptions as in theorem 3.1 except condition (ii) is dropped, but with the assumption that x^* and z^* are unique (this would be true if x^* and z^* are nondegenerate points of LP and the dual of LP respectively), then* $\lim_{k\to\infty} x^k = x^*$ *and* $\lim_{k\to\infty} \pi^k = \pi^*$. *Moreover, condition (iii) does not restrict the choice of s^k.*

As in the primal method, it is of practical importance to allow approximate solutions of the subproblem. A complication not present in the primal case is that the primal-variable estimates $x_j = w_j/(z_j + s_j)$ defined by (3.2) satisfy $Ax = b$ only in the limit. We therefore must introduce another primal estimate $\bar{x}$ that is "close" to x and satisfies $A\bar{x} = b$; see Section 5 for a discussion of how $\bar{x}$ may be computed in a Newton-based method.

Theorem 3.2. *(Convergence with approximate solutions of DSBP.) Let, π^k be an approximate solution of DSBP(w^k, s^k). For each π^k, the vectors z^k and x^k are defined by*

$$z^k = c - A^T\pi^k \quad \text{and} \quad x_j^k = \frac{w_j^k}{z_j^k + s_j^k}.$$

Let $\bar{x}^k$ denote any vector such that $A\bar{x}^k = b$. Assume the weights and shifts satisfy the conditions given in Theorem 3.1. Let β be a scalar such that $0 < \beta < 1$. If $\{w^k\}$ and $\{s^k\}$ also satisfy

$$w_j^k = \bar{x}_j^{k-1} s_j^k \quad \text{for } k \geq 1, \tag{3.7}$$

and if each member of the sequence $\{\pi^k\}$ satisfies

(i) $z^k + s^k > 0$;

(ii) $c^T\bar{x}^{k-1} - c^T\bar{x}^k \geq \beta \sum_{j=1}^{n} \frac{(z_j^k)^2 x_j^k}{s_j^k}$; *and*

(iii) $|a_j^T x^k - b_j| < \frac{s_j^k \bar{x}_j^{k-1}}{k!(z_j^k + s_j^k)}$,

then

$$\lim_{k\to\infty} \|x^{*k} - x^k\| = 0$$

and

$$\lim_{k\to\infty} \|\pi^{*k} - \pi^k\| = 0. \quad ∎$$

4. Properties of Shifted Barrier Subproblems

An unshifted barrier function such as (1.1) contains deliberately constructed singularities on the boundary of the feasible region. The higher-order terms of its Taylor expansion (those neglected in the quadratic model utilized by Newton's method) therefore become increasingly large *near* the boundary. When the solution of the original constrained problem lies on the boundary, the neighborhood in which each current model is "accurate" becomes smaller and smaller as the solution is approached. This property explains why Newton-based barrier methods may experience inefficiency when the starting point lies close to the boundary, since the *initial* Hessian is ill-conditioned (see below). With shifted barrier methods, a "good" choice of shifts can improve the condition of the initial Hessian matrix.

When the original linear programs are nondegenerate, the unshifted barrier subproblems do *not* suffer from a "traditional" flaw of barrier function methods, namely increasing ill-conditioning of the Hessian of the barrier function along the barrier trajectory. (See Gill et al.. [GMS+86] for a discussion.) Unfortunately, it is an "article of faith" that *practical linear programs are either primal or dual degenerate* (or both), which means that the Hessian of the barrier function does become singular in the limit.

4.1. Derivatives of the shifted primal barrier function

Let F_P denote the objective function of SBP(w, s):

$$F_P \equiv c^T x - \sum_{j=1}^{n} w_j \ln(x_j + s_j).$$

The gradient and Hessian matrix of F_P are given by

$$g_P = \nabla F_P = c - \sum_{j=1}^{n} \frac{w_j}{x_j + s_j} e_j \;\text{ and }\; H_P = \nabla^2 F_P = \operatorname{diag}\Big(\frac{w_j}{(x_j + s_j)^2}\Big). \tag{4.1}$$

Throughout this section, x^k denotes the solution of a primal shifted barrier subproblem SBP with weights w^k and shifts s^k.

Lemma 4.1. *(Bounded derivatives of the primal subproblem.) A sufficient condition that $\|H_P^k\| = \|H_P(x^k)\| < M < \infty$ is that $x_j^k + s_j^k \geq \epsilon > 0$.*

Proof. Let h_{jj} denote the j-th diagonal element of H_P^k. From (4.1) and the definition of z^k at the minimizer of the k-th subproblem, we have

$$h_{jj} = \frac{z_j^k}{x_j^k + s_j^k}.$$

The result follows directly from the boundedness of z_j^k and the assumption that $x_j^k + s_j^k \geq \epsilon > 0$. ∎

In the primal case, the barrier subproblem SBP has linear equality constraints. Assume without loss of generality that the columns of A are ordered so that A may be written as $\begin{pmatrix} B & S \end{pmatrix}$, where B is an $m \times m$ nonsingular matrix. The columns of the $n \times (n - m)$ matrix Z defined by

$$Z = \begin{pmatrix} -B^{-1}S \\ I \end{pmatrix} \tag{4.2}$$

form a basis for the null space of A, and the *reduced Hessian matrix* $Z^T H_P^k Z$ determines the condition of the solution of the subproblem.

Theorem 4.1. *(Bounded condition of the shifted primal subproblem.) Let $\{x^k\}$ denote the sequence defined in corollaries 2.1 or 2.2. Assuming the assumptions for those corollaries hold and $s_j^k \ge \theta\epsilon > 0$ when $z_j^k \ge \epsilon > 0$ then*

$$\|Z^T H_P^k Z\| \le \bar{M} < \infty$$

and

$$\text{cond}(Z^T H_P^k Z) \le \bar{M} < \infty.$$

Proof. The assumption that all dual solutions are nondegenerate implies x^* is unique and there exists at least $n - m$ elements of any z^* that are nonzero. There is no loss of generality if we assume $z_j^* \ge z_{\min}^* > 0$ for $j = m+1, \ldots, n$. It follows from (4.2) that

$$Z^T H_P^k Z = V = \text{diag}(v),$$

where $v_j = z_{m+j}^k/(x_{m+j}^k + s_{m+j}^k)$. Since $\lim_{k\to\infty} |z_i^k - z_i^{*k}| = 0$ it follows there exists $\epsilon > 0$ such that $z_{m+j}^k \ge \epsilon$ for $j = 1, \ldots, n-m$. From the assumption on s_i^k we have $v_{\min} \ge \epsilon/M > 0$, where $v_{\min}$ is the smallest element of v. We also have $v_{\max} \le z_{\max}/\theta\epsilon$, where $z_{\max}$ is the largest element of z_i^k for $i = m+1, \ldots, n$. It follows that

$$\|Z^T H_P^k Z\| \le \bar{M} < \infty$$

and

$$\text{cond}(Z^T H_P^k Z) \le M z_{\max}/\theta < \infty.$$

We have $\lim_{k\to\infty} \sup\{z_{\max}^k\} \le z_{\max}^*$ and $\lim_{k\to\infty} \inf\{z_i^k\} \ge z_{\min}^*$, where $z_{\max}^*$ is the largest element of z^*. The only restriction on θ is that $\theta\epsilon \le M$. It follows s^k may be chosen such that

$$\limsup_{k\to\infty} \text{cond}(Z^T H_P^k Z) \le z_{\max}^*/z_{\min}^*.$$

∎

4.2. Properties of the dual shifted barrier function

In considering the limiting behavior of dual barrier subproblems, we use π^k to denote the minimizer of the dual shifted objective function F_D of DSBP with weight and shift vectors w^k and s^k, and define z and x by (2.4) and (3.2).

Lemma 4.2. *(Bounded derivatives of the dual shifted subproblem.) A sufficient condition that $\|H_D^k\| = \|\nabla^2 F_D(\pi^k)\| < M < \infty$ is that $z_j^k + s_j^k \ge \epsilon > 0$.*

Proof. Using arguments similar to those of Lemma 4.1, we have

$$H_D^k = A(D^k)^2 A^T,$$

where $(D^k)^2 = \text{diag}(x_j^k/(z_j^k + s_j^k))$. The required result is immediate from the assumption $z_j^k + s_j^k \ge \epsilon > 0$. ∎

Since the dual barrier subproblem is unconstrained, the condition of its solution is determined by the Hessian matrix H_D^k.

Theorem 4.2. *(Bounded condition number of the dual subproblem.) Let $\{z^k\}$ denote the sequence defined in corollarys 3.1 or 3.2. Assuming the assumptions for those corollarys hold and $s_j^k \geq \theta\epsilon > 0$ when $x_j^k \geq \epsilon > 0$ then there exists a constant $\bar{M}$ such that*

$$\|H_D^k\| \leq \bar{M} < \infty,$$

$$\mathrm{cond}(H_D^k) \leq \bar{M} < \infty,$$

Proof. From the assumption that the primal solutions are nondegenerate x^* has at least m nonzero elements. We can assume without loss of generality that the nonzero elements of z^* occur in the first m positions. It follows from these remarks that z^* is unique and its first m elements of are nonzero. We have

$$H_D^k = BVB^T + CUC^T,$$

where B is the matrix composed of the first m columns of A, C the matrix composed of the remaining columns, $V = \mathrm{diag}(v)$ and $v_j = x_j^k/(z_j^k + s_j^k)$, $U = \mathrm{diag}(u)$ and $u_j = x_{j+m}^k/(z_{j+m}^k + s_{j+m}^k)$. The assumption of primal nondegeneracy means that B is of rank m. By a similar reasoning to that given in lemma 4.1 we can show V and U are bounded and that V has a bounded condition number.

∎

5. Newton's Method

In essentially all interior-point linear programming methods proposed to date, some version of Newton's method is applied to solve a barrier-type subproblem. The special features of the primal and dual problems imply that computation of the Newton iterate can be posed in terms of solving a least-squares problem. We now consider applying Newton's method to the subproblems SBP and DSBP arising in primal and dual shifted barrier methods.

5.1. Solving the primal subproblem

The shifted barrier subproblem SBP has a nonlinear objective function and linear equality constraints, and can be solved using various techniques (see, e.g., Gill, Murray and Wright [GMW81], Chapter 5). Let x denote the current estimate of the solution, where $x + s > 0$. At a typical iteration, the new estimate of the solution is $x + \alpha p$, where p is a search direction and α is a positive steplength.

In a projected Newton method for solving SBP (see Gill et al.. [GMS+86]), x satisfies $Ax = b$, and p is defined so that $A(x + \alpha p) = b$ for any step α, i.e., $Ap = 0$. The value of α is chosen to produce a "sufficient decrease" (in the sense of Ortega and Rheinboldt [OR70]) in the objective function of SBP, subject to the requirement that the new iterate remain strictly feasible with respect to the shifted bounds.

Given vectors x and π such that $Ax = b$ and $x + s > 0$, a typical projected Newton iteration proceeds as follows:

1. Solve the system
$$\begin{pmatrix} H_P & A^T \\ A & \end{pmatrix} \begin{pmatrix} -p \\ q \end{pmatrix} = \begin{pmatrix} g_L \\ 0 \end{pmatrix}, \tag{5.1}$$
where H_P is the Hessian of F_P, and g_L is given by $g_L = g_P - A^T\pi$ (see (2.5) and (4.1)).

2. Find a steplength α such that $x + \alpha p + s > 0$ and
$$|g_P(x + \alpha p)^T p| \le -\eta g_P(x)^T p,$$
where $0 \le \eta < 1$.

3. Set $x \leftarrow x + \alpha p$ and $\pi \leftarrow \pi + q$.

The vector g_L appears on the right-hand side of (5.1), so that the vector q is the *change* in the multiplier estimate.

The computational effort needed to perform this iteration is dominated by solving (5.1). As in [GMS+86], the diagonal structure of H_P can be used to transform (5.1) into a smaller system of order m. Let $\mu = \min_j w_j/(x_j + s_j)^2$ and write g_P and H_P as
$$g_P = c - \Delta e \;\text{ and }\; H_P = \mu D^{-2},$$
where $e = (1, 1, \ldots, 1)^T$, and Δ and D are the $n \times n$ diagonal matrices
$$\Delta = \operatorname{diag}\Big(\frac{w_j}{x_j + s_j}\Big) \;\text{ and }\; D = \operatorname{diag}(\sigma_j(x_j + s_j)),$$
where $\sigma_j = \sqrt{\mu/w_j}$. Substituting for g_L and H_P in (5.1) gives
$$\begin{pmatrix} \mu D^{-2} & A^T \\ A & \end{pmatrix} \begin{pmatrix} -p \\ q \end{pmatrix} = \begin{pmatrix} c - \Delta e - A^T\pi \\ 0 \end{pmatrix}, \tag{5.2}$$
from which it follows that p and q satisfy
$$AD^2A^Tq = AD^2(z - \Delta e), \;\text{ with }\; p = -\frac{1}{\mu}D^2(z - \Delta e - A^Tq). \tag{5.3}$$
Using the definition (2.3) of the Lagrangian, (5.3) may be written as
$$AD^2A^Tq = AD^2 g_L \;\text{ and }\; p = -\frac{1}{\mu}D^2(g_L - A^Tq),$$
which are the normal equations for the following weighted least-squares problem:
$$\underset{q \in \mathbb{R}^m}{\text{minimize}}\, \|D(g_L - A^Tq)\|_2^2. \tag{5.4}$$
(See Marxen [Mar89] for a derivation of analogous equations for the unshifted case.)

The scale factor μ is included in practice to ensure that the matrix D^2 is bounded when the shifts corresponding to zero components of x^* converge to a positive limit.

If the conditions of Lemma 2.1 hold and $\{s_j^k\}$ converges to a limit s_j^*, the j-th diagonal element of H_P converges to $z_j^* s_j^*/(x_j^* + s_j^*)^2$. We thus have

$$H_{jj} \to \begin{cases} z_j^*/s_j^* & \text{if } x_j^* = 0; \\ 0 & \text{if } x_j^* > 0, \end{cases} \tag{5.5}$$

which implies that at least m of the diagonal elements of H_P^{-1} become arbitrarily large.

The subproblem SBP involves minimization of the strictly convex function F_P within the subspace defined by $Ax = b$, and standard convergence theory for Newton's method with a linesearch can be applied. We simply state the conclusion that the Newton iterates will eventually converge to the unique solution of SBP(w, s).

Lemma 5.1. *Given the weight and shift vectors w and s, and an initial point x such that $Ax = b$, $x + s > 0$, the sequence of iterates generated by Newton's method as sketched above converges to $x^*(w, s)$.*

Proof. The proof follows from standard results described in Ortega and Rheinboldt [OR70]. ∎

In keeping with the aim of computing only an approximate solution of each subproblem, we now show that the conditions of Theorem 2.2 will eventually hold at a Newton iterate for the k-th subproblem.

Lemma 5.2. *(Characterization of Newton iterates for SBP.) Assume that the feasible region for problem LP is bounded, and let β be any scalar such that $0 < \beta < 1$. Let $\widehat{x}$ and $\widehat{\pi}$ satisfy $A\widehat{x} = b$, $\widehat{x} + s^k > 0$ and $\widehat{z} > 0$, where $\widehat{z} = c - A^T\widehat{\pi}$. If w^k and s^k satisfy*

$$w_j^k = \widehat{z}_j s_j^k \ \ \textit{and} \ \ s^k > 0, \tag{5.6}$$

and if Newton's method is applied to solve SBP(w^k, s^k) with starting values $(\widehat{x}, \widehat{z})$, then after a finite number of iterations, a point (x, π) will be reached such that

(i) $b^T\pi - b^T\widehat{\pi} \ge \beta \sum_{j=1}^{n} \dfrac{\widehat{z}_j x_j^2}{x_j + s_j^k}$; *and*

(ii) $|(g_L^k)_j| < \frac{1}{2} \dfrac{s_j^k \widehat{z}_j}{k!(x_j + s_j^k)}$.

Proof. For any x and π, rearranging expression (2.5) gives

$$z_j = (g_L^k)_j + \frac{w_j^k}{x_j + s_j^k}. \tag{5.7}$$

Since $\|g_L^k\|$ converges to zero as the Newton iterations proceed, the boundedness of the feasible region of LP implies that the dual has a nontrivial interior and

$z_j \to w_j^k/(x_j + s_j^k) > 0$, so that π is dual feasible. Moreover, since $(g_L^k)_j$ converges to zero (ii) must eventually be satisfied.

The same argument used in Theorem 2.1 shows that π and $\widehat{\pi}$ are related by the equation

$$b^T\pi - b^T\widehat{\pi} = x^T(\widehat{z} - z) \tag{5.8}$$

(see the analogous equation (2.11) of Theorem 2.1). Multiplying z by x^T (see (5.7)) and substituting the result in (5.8) gives

$$\begin{aligned} b^T\pi - b^T\widehat{\pi} &= \sum_{j=1}^{n} x_j\Big(\widehat{z}_j - \frac{\widehat{z}_j s_j^k}{x_j + s_j^k}\Big) - x^T g_L^k \\ &= \sum_{j=1}^{n} \frac{x_j^2 \widehat{z}_j}{x_j + s_j^k} - x^T g_L^k. \end{aligned} \tag{5.9}$$

Since $\|g_L^k\| \to 0$, a Newton iterate (x, π) must eventually be obtained such that

$$|x^T g_L^k| \le (1-\beta) \sum_{j=1}^{n} \frac{x_j^2 \widehat{z}_j}{x_j + s_j^k}.$$

The result follows immediately from (5.9). ∎

5.2. Solving the dual subproblem

The dual shifted barrier subproblem DSBP involves *unconstrained* minimization of the function

$$F_D(\pi) \equiv -b^T\pi - \sum_{j=1}^{n} w_j \ln(c_j - a_j^T\pi + s_j). \tag{5.10}$$

The gradient and Hessian matrix of F_D are given by

$$g_D = \nabla F_D = Ax - b \ \text{ and } \ H_D = \nabla^2 F_D = AD^2A^T,$$

where

$$D = \operatorname{diag}\Big(\frac{\sqrt{w_j}}{z_j + s_j}\Big) \ \text{ and } \ x_j = \frac{w_j}{z_j + s_j}, \tag{5.11}$$

In order to define a Newton method for solving DSBP, we assume that a point π is available such that $z + s > 0$ for the associated z (so that H_D is positive definite). Using the definitions above, the Newton search direction q satisfies

$$AD^2A^T q = Ax - b. \tag{5.12}$$

The function F_D is strictly convex, and is defined in a bounded region. As in the primal case, standard convergence theory (see, e.g., Ortega and Rheinboldt [OR70]) applies when Newton's method with a suitable linesearch (the obvious analogue of the primal method) is applied to minimize F_D.

Lemma 5.3. *Given weight and shift vectors w and s, let π be such that the vector x defined by (5.11) satisfies $Ax = b$ and $x + s > 0$. Consider applying a Newton method starting at π in which the search direction is defined by (5.12) and a standard linesearch is used to produce a sufficient decrease in F_D. Then the Newton iterates converge to $\pi^*(w, s)$, the unconstrained minimizer of $F_D(\pi)$.*

As in the primal case, the Newton iterates will eventually satisfy the conditions of Theorem 3.2.

Lemma 5.4. *(Characterization of Newton iterates for DSBP.) Assume that the feasible region for problem DLP is bounded. Let $\widehat{x}$ and $\widehat{\pi}$ be any vectors such that $\widehat{x} > 0$, $A\widehat{x} = b$ and $\widehat{z} + s^k > 0$, where $\widehat{z} = c - A^T\widehat{\pi}$. Assume that the weight and shift vectors w^k and s^k satisfy*

$$w_j^k = \widehat{x}_j s_j^k \quad \text{and} \quad s_j^k > 0. \tag{5.13}$$

For any iterate π of Newton's method applied to minimize F_D, we define z as $c - A^T\pi$, x from (5.11), and $\bar{x}$ as the closest point to x such that $A\bar{x} = b$. Then after a finite number of iterations, π and $\bar{x}$ will satisfy

(i) $c^T\widehat{x} - c^T\bar{x} \geq \beta \sum_{j=1}^{n} \dfrac{z_j^2 x_j}{s_j^k}$; *and*

(ii) $|a_j^T x - b_j| < \frac{1}{2} \dfrac{s_j^k \bar{x}_j}{k!(z_j + s_j^k)}$.

Proof. From the definition of x_j,

$$\bar{x}_j = x_j + (\bar{x}_j - x_j) = \frac{w_j^k}{z_j + s_j^k} + (\bar{x}_j - x_j). \tag{5.14}$$

Since the Newton iterates must eventually converge to $\pi^*(w, s)$ and the feasible region of DLP is bounded, we know that the primal has a nontrivial interior and $\bar{x}_j \to w_j^k/(z_j + s_j^k) > 0$. Consequently, since $\|Ax - b\| \to 0$ condition (ii) will eventually be satified.

Since $A\bar{x} = b$, we have

$$c^T\widehat{x} - c^T\bar{x} = z^T(\widehat{x} - \bar{x})$$

(*cf.* equation (3.6), Theorem 3.1). From the definition of x and w^k, x_j may be written in the form

$$x_j = \frac{w_j^k}{z_j + s_j^k} = \frac{\widehat{x}_j s_j^k}{z_j + s_j^k}. \tag{5.15}$$

Performing some simple rearrangement and substituting for x_j gives

$$\begin{aligned} c^T\widehat{x} - c^T\bar{x} = z^T(\widehat{x} - x - (\bar{x} - x)) &= \sum_{j=1}^{n} z_j(\widehat{x}_j - x_j) - z^T(\bar{x} - x) \\ &= \sum_{j=1}^{n} \frac{z_j^2 x_j}{s_j^k} - z^T(\bar{x} - x). \end{aligned} \tag{5.16}$$

Since $|\bar{x}_j - x_j| \to 0$, eventually we have

$$|z^T(\bar{x} - x)| \le (1 - \beta) \sum_{j=1}^{n} \frac{z_j^2 x_j}{s_j^k}. \tag{5.17}$$

The expression (ii) follows directly from (5.16) and (5.17). ∎

The proof of Theorem 5.4 holds if the primal-feasible vector $\bar{x}$ is chosen so that $\|\bar{x} - x\| \to 0$ as $k \to \infty$. Such a vector can be computed with little extra work once the Newton direction q is known. From (5.12) we have

$$A(x - D^2 A^T q) = b,$$

giving $\bar{x} = x - D^2 A^T q$. From the definition (5.11) of D^2, we have

$$\begin{aligned} \bar{x}_j &= \frac{w_j}{z_j + s_j} - \frac{w_j a_j^T q}{(z_j + s_j)^2} \\ &= w_j \frac{\bar{z}_j + s_j}{(z_j + s_j)^2}, \end{aligned} \tag{5.18}$$

where $\bar{z}_j$ is the j-th component of $\bar{z} = c - A^T(\pi + q)$.

6. Definition of the Shifts

In this section we briefly outline methods for initializing and updating the shifts. (The implementation of specific algorithms will be discussed in a future report.) We concentrate mainly on the primal method of Section 2, but much of the discussion applies directly to the dual method.

The convergence proofs of Sections 2 and 3 imply that there is a considerable degree of choice in the selection of the shifts as long as the weights are updated appropriately. The simplest strategy is thus to keep the shifts *constant* and update only the weights, so that

$$s_j^k = s_j^{k-1} \quad \text{and} \quad w_j^k = z_j^{k-1} s_j^k. \tag{6.1}$$

A disadvantage of this choice is that the sequence $\{x^k\}$ then converges at best linearly to x^*. To see why, assume that the conditions of Theorem 2.1 hold and that $\lim_{k\to\infty} z^k = z^*$. Let j denote the index of a variable that is strictly positive at the solution (i.e., $x_j^* > 0$). At the solution of the k-th subproblem we have $z_j^k = w_j^k/(x_j^k + s_j^k)$ (see (2.8)). Substituting for w_j^k from (6.1) gives

$$\frac{z_j^k}{z_j^{k-1}} = \frac{s_j^k}{x_j^k + s_j^k}. \tag{6.2}$$

If x_j^* is nonzero, z_j^* must be zero (see (1.5)), and (6.2) shows that z_j^k converges to zero at a linear rate.

A *superlinear* convergence rate may be obtained if s_j^k converges to zero for variables that are positive at the solution. For example, the relation

$$s_j^k = \frac{s_j^{k-1}}{x_j^{k-1} + s_j^{k-1}} \tag{6.3}$$

has the property that s_j^k converges to zero if $x_j^* > 0$ and to one if $x_j^* = 0$. In this case, w^k must converge z^* to retain the desired convergence properties.

Other schemes may be used to force the relevant elements of s^k to converge to zero at an even faster rate—e.g., the update $s_j^k = (s_j^{k-1})^2/(x_j^{k-1} + (s_j^{k-1})^2)$ retains many of the properties of (6.3). Other updates may be used to ensure that components converging to unity have a small deviation from unity compared to the difference between x^k and x^*; for example, we could define

$$s_j^k = \frac{(s_j^{k-1})^2}{(x_j^{k-1})^2 + (s_j^{k-1})^2}.$$

6.1. Initial values

In contrast to unshifted barrier methods, the shifted primal method does not require a "Phase 1" procedure to compute a strictly feasible initial point. Given any x that satisfies $Ax = b$, the shifts may be selected to make x interior with respect to the shifted bounds.

Given an arbitrary point x^0 at which initial weights and shifts are defined, the vector p such that $A(x^0 + p) = b$ satisfies

$$\begin{pmatrix} \mu D^{-2} & A^T \\ A & \end{pmatrix} \begin{pmatrix} -p \\ q \end{pmatrix} = \begin{pmatrix} g_L \\ r \end{pmatrix}, \tag{6.4}$$

where r is the residual vector $Ax^0 - b$. Using the techniques of Section 5, the search vector p of (6.4) can be computed from the equations

$$AD^2A^Tq = AD^2g_L + r \;\text{ and }\; p = -\frac{1}{\mu}D^2(g_L - A^Tq).$$

The weights and shifts are redefined at $x_0 + p$, and the iterations proceed as in Section 5.

The "optimal" amount of work to choose the initial weights and shifts depends on how much effort the subproblem is expected to require—i.e., how long the parameters will remain fixed. If the problem is well scaled, a possible choice for the initial shifts is

$$s_j^0 = \begin{cases} 1 & \text{if } x_j^0 \geq 0; \\ 1 - x_j^0 & \text{otherwise.} \end{cases} \tag{6.5}$$

With this strategy, $1/(s_j^0 + x_j^0) \leq 1$.

Given an initial estimate π^0, the initial weights should be chosen as

$$w_j^0 = z_j^0(s_j^0 + x_j^0),$$

where $z_j^0 = c_j - a_j^T \pi^0 > 0$. If no estimate of π is available, z_j^0 may be taken as either one or $s_j^0 + x_j^0$.

In all the definitions given above, it may be necessary to redefine the j-th shift if the j-th variable is not sufficiently positive.

6.2. Primal and dual methods based on Lagrangian shifts

A consistent theme in this paper has been that the limiting shifts should be strictly positive for variables that are zero at the solution, and zero for variables that are positive. With some safeguards, the vector z satisfies these requirements, and a possible choice for the shifts in the primal algorithm is

$$s_j^k = \mu z_j^{k-1}, \tag{6.6}$$

where z_j^{k-1} and $M > \mu > 0$ denotes the final reduced cost $c_j - a_j^T \pi^{k-1}$ of the previous subproblem. The parameter μ enables the initial shifts to be chosen so that $x^{-1} + s^0 > 0$ when x^{-1} and z^{-1} are prescribed initial estimate of x^* and z^* respectively. We shall refer to the value (6.6) as the *Lagrangian shift*, and a *primal Lagrangian shifted barrier method* involves a k-th subproblem of the form

$$\text{LSBP} \qquad \begin{array}{ll} \underset{x}{\text{minimize}} & c^T x - \displaystyle\sum_{j=1}^{n} \mu (z_j^{k-1})^2 \ln(x_j + \mu z_j^{k-1}) \\ \text{subject to} & Ax = b. \end{array} \tag{6.7}$$

In order to retain feasibility, the shift must be redefined if $x^{k-1} + z^{k-1} \leq 0$. If the dual problem is not degenerate, $x^k + z^k$ will eventually be positive. However, if both z_j and x_j are converging to zero, the shift may be repeatedly redefined as the solution is approached. In this case, the shift should be selected so that $z_j s_j/(x_j + s_j)^2$ is bounded away from zero, which implies that the condition of the reduced Hessian of the primal subproblem is bounded irrespective of dual nondegeneracy (see Theorem 4.1). Strategies for this case will be considered in a future paper.

The k-th Lagrangian dual shifted barrier subproblem is

$$\text{LDSBP} \qquad \underset{\pi}{\text{minimize}} \quad -b^T \pi - \sum_{j=1}^{n} \mu (x_j^{k-1})^2 \ln(c_j - a_j^T \pi + \mu x_j^{k-1}). \tag{6.8}$$

where $Ax^{k-1} = b$ and $x^{k-1} + z^{k-1} > 0$.

Appendix

Before presenting the main convergence results we prove three lemmas. In these lemmas $\{x^k\}$ and $\{s^k\}$ are sequences of scalars. In the main theorems they will be applied to the elements of x^k the similarity of notation is retained to make the use of the lemmas transparent.

Given a finite sequence $x^1, x^2, \ldots, x^k$ and a positive scalar ϵ_k, let $\mathcal{P}_k$ and $\mathcal{N}_k$ denote the index sets

$$\mathcal{P}_k = \{i \mid x^i \geq 0, \quad i \leq k\} \text{ and } \mathcal{N}_k = \{i \mid x^i \leq -M\epsilon_k^{1/2}, \quad i \leq k\}$$

respectively. We shall use the notation $|\mathcal{I}|$ to denote the number of elements in an index set $\mathcal{I}$.

Lemma 6.1. *Let $\{x^k\}$ and $\{s^k\}$ be two sequences of scalars. The sequence $\{s^k\}$ is such that $x^k + s^k > 0$, $0 < s^k < M$ and $s^j \leq M s^k$, $k < j$. Define z^k to be*

$$z^k = \frac{s^k}{x^k + s^k} z^{k-1},$$

where $z^0 > 0$. If the sequence $\{x^k\}$ is such that

(a) $\limsup_{k\to\infty} x^k = 0$;

(b) $\liminf_{k\to\infty} x^k < 0$; *and*

(c) $\lim_{k\to\infty} x^k z^k = 0$,

there exists a sequence $\{\epsilon_k\}$ such that $\lim_{k\to\infty} \epsilon_k = 0$ and $\lim_{k\to\infty} |\mathcal{N}_k|/|\mathcal{P}_k| = 0$.

Proof. From (b), there exists a positive scalar σ and a subsequence $\{x^{k_i}\}$ such that $x^{k_i} < -\sigma$. It follows from (c) that $\lim_{i\to\infty} z^{k_i} = 0$ and so there must be a subsequence $\{z^{l_i}\}$ such that

$$z^{l_i} < z^{l_{i-1}}.$$

Let $\mathcal{J}^i$ denote the set of consecutive indices $l_{i-1}, \ldots, l_i$. From the definition of z^k we have

$$z^{l_i} = \prod_{j\in\mathcal{J}^i} \theta_j z^{l_{i-1}}, \text{ where } \theta_j = s^j/(x^j + s^j).$$

It follows that

$$\prod_{j\in\mathcal{J}^i} \theta_j < 1.$$

Define

$$\sigma_i = \max\{x^j \mid j \in \mathcal{J}^i\} \text{ and } \beta_i = \sup\{x^j \mid j \geq l_{i-1}\}$$

and we extend the sequence $\{\epsilon_i\}$ for all k by defining $\epsilon_k = \beta_i$, where l_{i-1} is the smallest index in $\mathcal{L}$ such that $i \geq k$. It follows from these definitions and (a) that $\lim_{k\to\infty} \epsilon_k = 0$.

The indices of $\mathcal{J}$ may be partitioned into disjoint subsets $\mathcal{J}_+$, $\mathcal{J}_-$ and $\mathcal{J}_0$, where $\mathcal{J}_+ = \{j \mid x^j \geq 0\}$ $\mathcal{J}_- = \{j \mid x^j < -M\epsilon_j^{1/2}\}$. It follows that

$$\prod_{j \in J_+} \theta_j \prod_{j \in J_-} \theta_j \prod_{j \in J_0} \theta_j < 1. \tag{6.9}$$

For indices $j \in \mathcal{J}_0$,

$$\theta_j = \frac{1}{1 + x_j/s_j} > 1.$$

For indices $j \in J_+$

$$\theta_j \geq 1/(1 + \epsilon_j/s^j).$$

and $-\sigma < x_j \leq \sigma_j \leq \beta_j$ (?). It follows from $Ms^k \geq s^{k-1}$, $x^k + s^k > 0$ and condition (b) that $s^k > \sigma/M > 0$, for all k. Consequently, for $j \in J_+$

$$\theta_j \geq 1/(1 + Mx^j/\sigma) \geq 1/(1 + M\beta_i/\sigma). \tag{6.10}$$

Finally, for indices $j \in J_-$

$$\theta_j \geq 1/(1 - M\epsilon_j^{1/2}/s^j) = 1/(1 - M\beta_i/s^j).$$

It follows for $j \in J_-$ that

$$\theta_j \geq 1/(1 - \beta_i^{1/2}). \tag{6.11}$$

From (6.9), (6.10) and (6.11) we get

$$(1 - \beta_i^{1/2})^{|\mathcal{J}_-|}(1 + M\beta_i/\sigma)^{|\mathcal{J}_+|} > 1.$$

Since $\sigma > 0$ and $\lim_{i\to\infty} \beta_i = 0$ it follows that

$$\lim_{i\to\infty} |\mathcal{J}_-^i|/|\mathcal{J}_+^i| = \lim_{i\to\infty} O(\beta_i^{1/2}) = 0.$$

Since $|\mathcal{P}_{l_i}| = \sum_{j=1}^{i} |\mathcal{J}_+^j|$ and $|\mathcal{N}_{l_i}| = \sum_{j=1}^{i} |\mathcal{J}_-|$ the required results follows. ∎

Lemma 6.2. *If the relationship*

$$z^k = \frac{s^k}{x^k + s^k} z^{k-1}$$

in Lemma 6.1 is replaced by

$$z^k = \frac{\psi_k s^k}{x^k + s^k} z^{k-1},$$

where $\prod_{k=1}^{\infty} \psi_k = \gamma$ *and* $0 < \epsilon \leq \gamma \leq M < \infty$ *, then the lemma still holds.*

Proof. The subsequence $\{l_i\}$ could be chosen such that

$$z^{l_i} < \gamma z^{l_{i-1}}.$$

Consequently, we still require

$$\prod_{j=1}^{L_i} \theta_j < 1.$$

∎

Lemma 6.3. *Let $\{x_j^k\}$, $\{z_j^k\}$ and $\{s_j^k\}$ for $j = 1, \ldots, n$ satisfying the properties of the sequences $\{x^k\}$, $\{z^k\}$ and $\{s^k\}$ defined in Lemma 6.1 or Lemma 6.2. If P_k is the number of the first k elements of these sequences that satisfy the n inequalities: $x_j^i \geq -M(\epsilon_k^i)^{1/2}$, for $i = 1, \ldots, n$, then*

$$\lim_{k\to\infty} \frac{k}{P_k} = 1.$$

Proof. Assume

$$\lim_{k\to\infty} \frac{k}{P_k} < 1.$$

Let N_k denote the number of times $x_n^i < -M(\epsilon_n^i)^{1/2}$, $i = 1, \ldots, k$. There is no loss of generality if we assume the sequence that satisfies $x_j^i < -M\epsilon_j^i$ the most is $\{x_n^j\}$. It follows that

$$N_k \geq (k - P_k)/n.$$

Therefore,

$$\lim_{k\to\infty} \frac{N_k}{k} \geq \lim_{k\to\infty} \frac{1}{n}(1 - P_k/k) > 0.$$

This inequality contradicts lemma 6.1. ∎

References

[Fia79] A. V. Fiacco. Barrier methods for nonlinear programming. In A. Holzman, editor, *Operations Research Support Methodology*, pages 377–440, New York, 1979. Marcel Dekker. 2

[Fle81] Roger Fletcher. *Practical Methods of Optimization.* Volume 2: Constrained Optimization. John Wiley and Sons, Chichester and New York, 1981. 2

[FM68] Anthony V. Fiacco and Garth P. McCormick. *Nonlinear Programming: Sequential Unconstrained Minimization Techniques.* John Wiley and Sons, Inc., New York-London-Sydney, 1968. 2

[Fri55] K. R. Frisch. The logarithmic potential method of convex programming. Memorandum of May 13, University Institute of Economics, Oslo, Norway, 1955. 2

[GMS+86] Philip E. Gill, Walter Murray, Michael A. Saunders, J. Tomlin, and Margaret H. Wright. On projected Newton methods for linear programming and an equivalence to Karmarkar's projective method. *Math. Program.*, 36, 183–209, 1986. 16, 18, 19

[GMW81] Philip E. Gill, Walter Murray, and Margaret H. Wright. *Practical Optimization*. Academic Press Inc. [Harcourt Brace Jovanovich Publishers], London, 1981. 2, 18

[Kar84] N. Karmarkar. A new polynomial-time algorithm for linear programming. *Combinatorica*, 4, 373–395, 1984. 3

[Mar89] A. Marxen. Primal barrier methods for linear programming. Report SOL 89-6, Department of Operations Research, Stanford University, Stanford, CA, 1989. 19

[OR70] J. M. Ortega and W. C. Rheinboldt. *Iterative solution of nonlinear equations in several variables*. Academic Press, New York, 1970. 18, 20, 21

[Osb72] M. R. Osborne. Topics in optimization. Report CS 72–279, Department of Computer Science, Stanford University, Stanford, CA, 1972. 2

[Pow69] Michael J. D. Powell. A method for nonlinear constraints in minimization problems. In Roger Fletcher, editor, *Optimization*, pages 283–298, London and New York, 1969. Academic Press. 2

[Ste88] G. W. Stewart. On scaled projections and pseudo-inverses. Report TR-89, Department of Computer Science, University of Maryland, College Park, MD, 1988. 4